

# A Note on the High-dimensional Sparse Fourier Transform in the Continuous Setting

Liang Chen\*

## Abstract

In this paper, we theoretically propose a new hashing scheme to establish the sparse Fourier transform in high-dimensional space. The estimation of the algorithm complexity shows that this sparse Fourier transform can overcome the curse of dimensionality. To the best of our knowledge, this is the first polynomial-time algorithm to recover the high-dimensional continuous frequencies.

**Keywords:** Curse of dimensionality, Frequency estimation, Runtime complexity, Sparse Fourier transform.

## 1 Introduction

Discrete Fourier transform (DFT) plays a fundamental role in signal processing, scientific computing and other fields. Its runtime complexity can reach  $N\log(N)$  (where  $N$  represents the signal size) by the fast Fourier transform algorithm. This runtime complexity is still too high for the ultrawide bandwidth signals or high-dimensional signals. If we increase the prior information of the signals, that is, assuming the sparsity of the frequencies, then the sparse discrete Fourier transform (SFT) [1] will have more advantages than the DFT. The sample complexity and runtime complexity of the SFT are mainly affected by the sparsity, and less affected by the bandwidth. The SFT originated from the work on the Hadamard transform [2] and has received continuous attention from applied mathematics [3, 4, 5, 6, 7, 8], signal processing [9, 10, 11, 12, 13, 14], and theoretical computer science communities [15, 16, 17, 18, 19, 20] over the last two decades. Most of the relevant works deals with the discrete case where the frequencies are on the grid. Under such condition, the SFT can overcome the curse of dimensionality [20, 21, 22, 23]. However, this condition that the frequencies are on the grid is so strong. It is natural for people to consider the case that the frequencies are in a continuous region. This has led researchers to establish the sparse Fourier transform in the one-dimensional continuous setting [18, 24, 25].

Recently, [26] initiates the study on the SFT in the high-dimensional continuous setting. Unfortunately, their SFT method is still subject to the curse of dimensionality, namely, its runtime complexity is greater than  $2^{O(d)}$  ( $d$  stands for the dimension). The main difficulty in estimating high-dimensional continuous frequencies is that when we fix a set of discrete orthogonal bases whose frequencies are located on the grid, there is always a single-frequency signal with the frequency located in a continuous region, which is not sparse under the discrete Fourier transform (see “the windowing effect” in [11]),

---

\*Department of Mathematics, Jiujiang University, Jiujiang, 332000, China. E-mail address: *chen-liang3@mail2.sysu.edu.cn*. Supported in part by National Natural Science Foundation of China under grant 11971490, and by Natural Science Foundation of Guangdong Province under grant 2018A030313841.



and as the dimension increases, the  $L_1$  norm of the frequency coefficients increases exponentially. On the other hand, in a physical sense, the single-frequency signals should obviously be sparse.

In this paper, we present a new hashing scheme to transform the high-dimensional SFT into the one-dimensional SFT. The computational complexity of this algorithm is polynomial, which means that the algorithm can break the curse of dimensionality. To the best of our knowledge, this is the first polynomial-time algorithm to recover the off-grid high-dimensional frequencies.

Formally, we consider the signal  $f$  of the following form

$$f(t) = \sum_{j=1}^k f_j(t) \triangleq \sum_{j=1}^k a_j \exp(2\pi i w_j \cdot t), \quad (1.1)$$

where  $t \in \mathbb{R}^d$ ,  $w_j \in [-M, M]^d$ ,  $a_j \in \mathbb{C}$ ,  $0 < A' \leq |a_j| \leq A$  for all  $j = 1, 2, \dots, k$  and  $\min_{1 \leq i < j \leq k} |w_i - w_j| > \eta$ .

Our goal is to recover  $w_j, a_j, j = 1, 2, \dots, k$ .

## 2 Main Result

We first give some symbols and notions. Let  $\text{sgn}(x) = 1$ , when  $x \geq 0$ ; otherwise,  $\text{sgn}(x) = -1$ . Let  $J^*$  denote the transpose of the matrix  $J$ . Denote for each  $t \in \mathbb{R}$  by  $\lfloor t \rfloor$  the largest integer not bigger than  $t$ . Let us say that  $U \subset E$  is the  $\epsilon$ -support set of the function  $\hat{f}$  if  $\int_{E \setminus U} |\hat{f}|^2 \leq \epsilon^2$ , where  $E$  is the domain of  $\hat{f}$ . Let

$$\Gamma_1 \triangleq \left\{ \left( \frac{x_1}{F}, \dots, \frac{x_d}{F} \right) : x_i \in \mathbb{Z} \cap \left[ \frac{-TF}{2}, \frac{TF}{2} \right], 1 \leq i \leq d \right\},$$

and

$$\Gamma_2 \triangleq \left\{ \left( \frac{\xi_1}{T}, \dots, \frac{\xi_d}{T} \right) : \xi_i \in \mathbb{Z} \cap \left[ \frac{-TF}{2}, \frac{TF}{2} \right], 1 \leq i \leq d \right\}.$$

The discrete Fourier transform of the function  $g$  takes the following form

$$\mathcal{F}[g](\xi) = \frac{1}{(\sqrt{TF})^d} \sum_{x \in \Gamma_1} g(x) \exp(-2\pi i x \cdot \xi), \quad \xi \in \Gamma_2,$$

and the inverse discrete Fourier transform of the function  $g$  is defined by

$$\mathcal{F}^{-1}[g](x) = \frac{1}{(\sqrt{T})^d} \sum_{\xi \in \Gamma_2} g(\xi) \exp(2\pi i x \cdot \xi), \quad x \in \Gamma_1.$$

For  $x, y \in \Gamma_1$ ,  $x \pm y \triangleq x \pm y(\text{mod } T) \in \Gamma_1$ . For  $x, y \in \Gamma_2$ ,  $x \pm y \triangleq x \pm y(\text{mod } F) \in \Gamma_2$ .

The “bucket” in the frequency domain is defined by

$$B_j \triangleq \left\{ (\xi_1/T, \dots, \xi_d/T) \in \Gamma_2 : \xi_d \in \left[ \frac{TF(j-1)}{s} - \frac{FT}{2}, \frac{TFj}{s} - \frac{FT}{2} \right) \right\}, \quad 1 \leq j \leq s. \quad (2.2)$$

The hashing transform (in the frequency domain) is defined by  $H(\xi) \triangleq h\xi - (0, \dots, 0, b/T)^*$ , where

$$h = \begin{bmatrix} \mathbf{I}_{d-1} & 0 \\ V_h & h_d \end{bmatrix}, \quad (2.3)$$



here  $\mathbf{I}_{d-1}$  is the identity matrix of order  $d-1$ ,  $V_h = (h_1, \dots, h_{d-1})$  and  $\{h_1, \dots, h_d\}$  are independent draws from the uniform distribution on the set  $\{2n+1 : n \in \mathbb{Z}\} \cap [-\sqrt{d}F/(\eta), \sqrt{d}F/(\eta)]$ . The random variable  $b$  obeys the uniform distribution on the set  $\mathbb{Z} \cap \{[-\frac{TF}{2s}, \frac{TF}{2s})\}$ .

In this paper, we assume that  $T, F, s$  are powers of 2 and  $1 < s < F$ . Next, we give a key lemma.

**Lemma 2.1** *Suppose  $0 < \delta < \frac{1}{2}$ ,  $0 < \epsilon < \min\{1, \eta, \frac{A'}{4}, \frac{1}{4A^2}\}$ ,  $s = \mathcal{O}(\frac{k^2}{\delta})$ ,  $T = \mathcal{O}(k^5 s d^{7/2} / (\epsilon^2 \eta \delta^3))$ ,  $F = sM$ . With probability at least  $1 - \delta/4$  over the randomness of  $\{h_1, \dots, h_d, b\}$ , for each  $j \in \{1, \dots, s\}$ , either*

$$I_{j,1} \triangleq \frac{1}{(TF)^d} \sum_{x \in \Gamma_1} |e^{\frac{2\pi i b (h_d)^{-1} x_d}{TF}} \mathcal{F}^{-1}[\mathcal{X}_{B_j} \cdot \mathcal{F}[f_H]]((h^{-1})^* x)|^2 \leq A^2 \epsilon^2 \delta / (ds). \quad (2.4)$$

or there exists a unique  $j_k \in \{1, 2, \dots, k\}$  such that

$$I_{j,2} \triangleq \frac{1}{(TF)^d} \sum_{x \in \Gamma_1} |f_{j_k}(x) - e^{\frac{2\pi i b (h_d)^{-1} x_d}{TF}} \mathcal{F}^{-1}[\mathcal{X}_{B_j} \cdot \mathcal{F}[f_H]]((h^{-1})^* x)|^2 \leq A^2 \epsilon^2 \delta / (ds). \quad (2.5)$$

Where

$$f_H(x) \triangleq f(h^* x) \exp(-\frac{2\pi i b x_d}{TF}), \quad (2.6)$$

and

$$\mathcal{X}_{B_j}(\xi) \triangleq \begin{cases} 1 & \xi \in B_j \\ 0 & \xi \notin B_j \end{cases}, \quad j = 1, 2, \dots, s.$$

By Markov's inequality, we have a corollary.

**Corollary 2.2** *All parameters are set as Lemma 2.1. For any  $l \in \{1, 2, \dots, d\}$ , choosing the random vector  $X_l \triangleq (x_1, \dots, x_{l-1}, x_{l+1}, \dots, x_d)$  which obeys the uniform distribution on the set  $\mathbb{Z}^{d-1} \cap [-\frac{TF}{2}, \frac{TF}{2})^{d-1}$ , then with probability at least  $1 - \delta/(4ds)$  over the randomness of  $(x_1, \dots, x_{l-1}, x_{l+1}, \dots, x_d)$ , we have*

$$\sum_{x_l \in \mathbb{Z} \cap [-\frac{TF}{2}, \frac{TF}{2})} \frac{1}{TF} |e^{\frac{2\pi i b (h_d)^{-1} x_d}{TF}} \mathcal{F}^{-1}[\mathcal{X}_{B_j} \cdot \mathcal{F}[f_H]](h^*(x))|^2 \leq ds I_{j,1} / \delta$$

or

$$\sum_{x_l \in \mathbb{Z} \cap [-\frac{TF}{2}, \frac{TF}{2})} \frac{1}{TF} |f_{j_k}(x) - e^{\frac{2\pi i b (h_d)^{-1} x_d}{TF}} \mathcal{F}^{-1}[\mathcal{X}_{B_j} \cdot \mathcal{F}[f_H]](h^*(x))|^2 \leq ds I_{j,2} / \delta.$$

Where  $x = (x_1/F, \dots, x_{l-1}/F, x_l/F, x_{l+1}/F, \dots, x_d/F)^*$ .

We explain Lemma 2.1 at a high level. It is well known that for the discrete Fourier transform, enlarge the sampling interval can enhance the frequency concentration (the frequency resolution) of each component  $f_j$ , and the frequency gap between different components is greater than a constant value  $\eta$ . Therefore, with the expanding of the sampling interval,  $\eta$  will be greater than the radius  $\beta_\epsilon$  of the  $\epsilon$ -support set (let  $\frac{d^{3/2} A^2}{T \beta_\epsilon} = \epsilon^2 / k^2$  in Lemma 3.1) for each signal component in the frequency domain. And the hashing transform we defined can transform the gap between frequencies to the gap between the last coordinates of the frequencies (see Lemma 3.2). When  $\eta > \mathcal{O}(dk^3 \beta_\epsilon / \delta^2)$ , this hashing transform can keep the main frequencies  $\phi_{j, \beta_\epsilon}$  (defined in Lemma 3.1) of the same component into the same bucket, and different components'  $\epsilon$ -support sets  $\phi_{i, \beta_\epsilon}$ ,  $\phi_{j, \beta_\epsilon}$ ,  $i \neq j$  can be isolated into



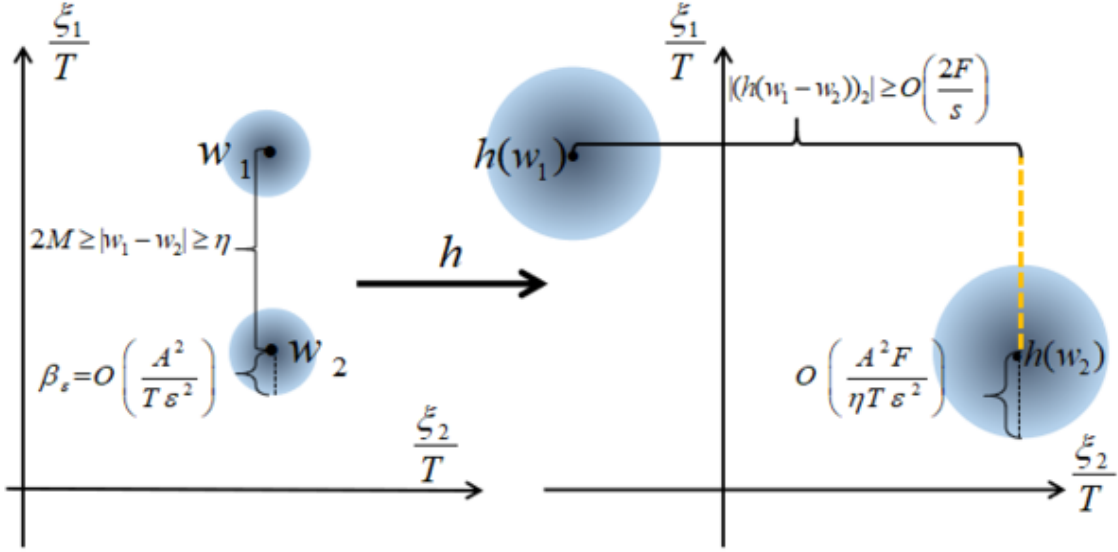


Figure 2.1: The left side of the figure shows the spectrogram of the two-dimensional signal  $\exp(i2\pi w_1 \cdot t) + \exp(i2\pi w_2 \cdot t)$  under the discrete Fourier transform, where  $\beta_\epsilon$  is the radius of the  $\epsilon$ -support set  $\phi_{1,\beta_\epsilon}, \phi_{2,\beta_\epsilon}$  in the frequency domain. The right side of the figure is the spectrogram after hashing transform, by Lemma 3.2, we have  $|(h(w_1 - w_2))_2| > 2F/s$ . By increasing the sampling interval  $T$ , we can make  $\mathcal{O}\left(\frac{A^2 F}{T\eta\epsilon^2}\right) < \mathcal{O}(F/s)$ .

different buckets (see Corollary 3.3). The figure 2.1 shows how the frequencies of the two-dimensional signal is transformed under the hashing operator (2.3).

When the different components of the signal are isolated, we only need to recover the frequency and amplitude in each bucket. By Lemma 2.1, there are only two cases of such amplitude and frequency, one is that they are close to the amplitude and frequency of a certain component (up to the hashing transformation), and the other is that the amplitude is less than  $\mathcal{O}(\epsilon)$ .

Specifically, for  $1 \leq j \leq s$ ,  $1 \leq l \leq d$ , set

$$g_{j,l}(t) = \exp\left(2\pi i b(h_d)^{-1} \rho_l\left(\frac{\lfloor tF \rfloor}{TF}\right)\right) \mathcal{F}^{-1}[\mathcal{X}_{B_j} \cdot \mathcal{F}[f_H]](h^* x_{l,t}),$$

where  $\rho_l\left(\frac{\lfloor tF \rfloor}{TF}\right) = \frac{\lfloor tF \rfloor}{TF}$ , when  $l = d$ ; otherwise,  $\rho_l\left(\frac{\lfloor tF \rfloor}{TF}\right) = \frac{x_d}{TF}$  and

$$x_{l,t} = (x_1, \dots, x_{l-1}, \lfloor tF \rfloor / F, x_{l+1}, \dots, x_d).$$

Here  $(x_1, \dots, x_{l-1}, x_{l+1}, \dots, x_d)$  is the random vector obeys the uniform distribution on the set  $\mathbb{Z}^{d-1} \cap [-\frac{TF}{2}, \frac{TF}{2}]^{d-1}$ . The sample value of the function  $g_{j,l}$  is calculated by the method in Lemma 4.1. Let  $F \geq \mathcal{O}(\sqrt{d}M/\epsilon)$ , by Lemma 2.1 and Corollary 2.2, either there is a signal component  $f_{j_k}$  with frequency  $w_{j_k}$  such that

$$\frac{1}{T} \int_{-T/2 < t < T/2} |g_{j,l}(t) - a_{j_k} \theta_{j_k,l} e^{2\pi i w_{j_k,l} t}|^2 dt \leq \mathcal{O}(A^2 \epsilon^2),$$



or

$$\frac{1}{T} \int_{-T/2 < t < T/2} |g_{j,l}(t)|^2 dt \leq \mathcal{O}(A^2 \epsilon^2),$$

where  $w_{j_k,l}$  denotes  $l$ -th component of the vector  $w_{j_k}$  and

$$\theta_{j_k,l} = \exp \left( \frac{2\pi i}{F} (w_{j_k,1}x_1 + \dots + w_{j_k,l-1}x_{l-1} + w_{j_k,l+1}x_{l+1} + \dots + w_{j_k,d}x_d) \right).$$

We have the following *algorithm*: For each  $j \in \{1, 2, \dots, s\}$ , we first use the algorithm in [24] to recover the amplitude and frequency of  $g_{j,1}(t)$ . If the absolute value of the output amplitude less than  $A'/2$ , we turn to recover the amplitude and frequency of  $g_{j+1,1}(t)$ ; otherwise, we keep  $w_{j_k,1}^o$  and continue to use the algorithm in [24] to recover the frequencies of  $g_{j,2}(t), \dots, g_{j,d}(t)$ . Finally,  $\{w_{j_k,1}^o, \dots, w_{j_k,l}^o, \dots, w_{j_k,d}^o\}$  can be recovered element-wisely.

*Note*: From Theorem 1.1 in [24], it takes at most  $\mathcal{O}(ds \log(TF) \log(ds/(\delta\epsilon))/\delta)$  samples of the function  $g_{j,l}$  to recover the frequency  $w_{j_k,l}$  (up to the accuracy  $A^2\epsilon/(TA')$ ) with probability at least  $1 - \delta/(4ds)$ . By Lemma 4.1, in order to get  $\mathcal{O}(ds \log(TF) \log(ds/(\delta\epsilon)))$  samples (up to the accuracy  $A^2\epsilon$ ) of  $g_{j,l}$  with probability at least  $1 - \delta$ , we need a total of

$$\mathcal{O}(k^2 ds \ln^2(TF + 1) \log(TF) \log(TF ds/(\delta\epsilon)) \ln^2(1/\delta)/(\delta\epsilon^2))$$

samples and running time. Therefore, we have the following result.

**Theorem 2.3** All parameters are set as Lemma 2.1, besides,  $F \geq \mathcal{O}(\sqrt{d}M/\epsilon)$ . The above algorithm can output  $\{w_1^o, \dots, w_k^o\}$  such that

$$|w_j^o - w_j| < \mathcal{O}\left(\frac{\epsilon A^2 \sqrt{d}}{A'T}\right) < \mathcal{O}\left(\frac{A^2 \epsilon^3 / A'}{\sqrt{d}k^4}\right), \quad i = 1, 2, \dots, k.$$

holds with probability at least  $1 - \delta$  over the randomness of  $\{h_1, \dots, h_d, b\}$ ,  $\{X_l\}_{l=1}^d$  (see Corollary 2.2),  $\mathcal{T}$  (see Lemma 4.1) and the algorithm in [24]. The runtime complexity and sample complexity are at most

$$\mathcal{O}(k^2 d^2 s^2 \ln^2(TF + 1) \log(TF) \log(TF ds/(\delta\epsilon)) \ln^2(1/\delta)/(\epsilon^2 \delta)).$$

The success probability can be boosted to 1 by repeatedly restarting (as indicated in ([15, 24])).

Next, we consider restoring the amplitudes. Notice that  $|w_j^o - w_j| < \mathcal{O}(\frac{A^2 \epsilon^3 / A'}{\sqrt{d}k^4})$  and  $\epsilon < \eta$ . As the sampling interval increases, the different signal components  $f_j$  are nearly orthogonal, that is

$$\begin{aligned} & (\epsilon^2/k)^d \int_{t \in [0, k/\epsilon^2]^d} f(t) e^{-2\pi i w_j^o \cdot t} dt \\ &= a_j (\epsilon^2/k)^d \int_{t \in [0, k/\epsilon^2]^d} e^{2\pi i (w_j - w_j^o) \cdot t} dt + \sum_{l \neq 1}^k (\epsilon^2/k)^d \int_{t \in [0, k/\epsilon^2]^d} a_l e^{2\pi i (w_l - w_j^o) \cdot t} dt \\ &= a_j (1 + \mathcal{O}(\frac{A^2 \epsilon / A'}{k^3})) + \sum_{l \neq 1}^k (\epsilon^2/k)^d \int_{t \in [0, k/\epsilon^2]^d} a_l e^{2\pi i (w_l - w_j^o) \cdot t} dt \\ &\leq a_j (1 + \mathcal{O}(\frac{A^2 \epsilon / A'}{k^3})) + (k-1) A \mathcal{O}(\epsilon^2 / \eta k). \end{aligned} \tag{2.7}$$

Then

$$a_j = (\epsilon^2/k)^d \int_{t \in [0, k/\epsilon^2]^d} f(t) e^{-2\pi i w_j^o \cdot t} dt + \mathcal{O}(A^2 \epsilon / (A')). \tag{2.8}$$

We can use the Monte Carlo method to compute the integral

$$(\epsilon^2/k)^d \int_{t \in [0, k/\epsilon^2]^d} f(t) e^{-2\pi i w_j^o \cdot t} dt.$$



Choosing the random samples  $\{f(t_i)e^{-2\pi i w_j^o \cdot t_i}\}_{i=1}^{N_\epsilon}$  obeying the uniform distribution on  $[0, k/\epsilon^2]^d$ , where  $N_\epsilon = \mathcal{O}(\log(1/\delta)A'^2/(A\epsilon)^2)$ . By Hoeffding's inequality, with probability at least  $1 - \delta$ ,

$$\left| \frac{1}{N_\epsilon} \sum_{i=1}^{N_\epsilon} f(t_i)e^{-2\pi i w_j^o \cdot t_i} - (\epsilon^2/k)^d \int_{t \in [0, k/\epsilon^2]^d} f(t)e^{-2\pi i w_j^o \cdot t} dt \right| < A^2\epsilon/A' \quad (2.9)$$

Therefore, the amplitude  $a_j$  can be recovered (up to  $\mathcal{O}(A^2\epsilon/A')$ ).

### 3 Proof of Lemma 2.1

To prove Lemma 2.1, we need some technical lemmas.

**Lemma 3.1** *Let  $0 < \beta < F/2$ , for each  $f_j$ , the following inequalities*

$$\frac{1}{T^d} \sum_{\xi \in \Gamma_2} |\mathcal{X}_{\phi_{j,\beta}}(\xi) \cdot \mathcal{F}[f_j](\xi) - \mathcal{F}(f_j)(\xi)|^2 \leq \mathcal{O}\left(\frac{d^{3/2}A^2}{T\beta}\right), \quad (3.10)$$

*holds for any  $j \in \{1, 2, \dots, k\}$ . Where*

$$\mathcal{X}_{\phi_{j,\beta}}(\xi) \triangleq \begin{cases} 1 & \xi \in \phi_{j,\beta} \triangleq \{\xi : |\xi - w_j| \leq \beta/2\} \\ 0 & \text{otherwise} \end{cases}, j = 1, \dots, k.$$

*Proof:* For any  $\alpha \in (-1/T, 1/T)$  and any  $n \in \mathbb{Z} \cap ([2, FT/2) \cup [-FT/2, -2])$ , we have

$$\frac{1}{TF} \left| \sum_{j \in \mathbb{Z} \cap [-\frac{TF}{2}, \frac{TF}{2}]} \exp(2\pi i \alpha j / F) \exp\left(\left(-\frac{2\pi i j}{F}\right)\left(\frac{n}{T}\right)\right) \right| \leq \frac{2}{TF|1 - \exp(2\pi i (1/F)(\alpha - \frac{n}{T}))|} \leq \mathcal{O}\left(\frac{1}{n}\right). \quad (3.11)$$

Since  $|a_j| \leq A$ , by (3.11), using Parseval's identity, we have

$$\frac{1}{T^d} \sum_{\xi \in \Gamma_2} |\mathcal{X}_{\phi_{j,\beta}}(\xi) \cdot \mathcal{F}[f_j](\xi) - \mathcal{F}(f_j)(\xi)|^2 \leq \mathcal{O}\left(\frac{d^{3/2}A^2}{T\beta}\right), \quad (3.12)$$

which completes the proof.  $\square$

**Lemma 3.2** *Let  $s = \mathcal{O}(k^2/\delta)$ ,  $F = sM$ . For any vector  $w' \in [-2M, 2M]^d$  with  $|w'_d| \geq \eta$ , the inequality  $|(h(w'))_d| > 2F/s$  holds with probability at least  $1 - \delta/(8k^2)$  over the randomness of  $h_1, h_2, \dots, h_d$ , where  $(h(w'))_d$  denotes  $d$ -th component of the vector  $h(w')$ ,  $w'_i$  denotes  $i$ -th component of the vector  $w'$ ,  $\{h_1, \dots, h_d\}$  are independent draw from the uniform distribution on the set  $\{2n + 1 : n \in \mathbb{Z}\} \cap [-\sqrt{d}F/(\eta), \sqrt{d}F/(\eta)]$ .*

*Proof:* Without loss of generality, let us assume that the  $d$ -th component  $w'_d$  of  $w$  is greater than  $\eta/\sqrt{d}$ . Fix arbitrary  $h_1, \dots, h_{d-1}$ , since  $\eta/\sqrt{d} \leq |w'_d| \leq 2M$ , with probability at most  $\mathcal{O}(1/s)$  (no greater than the ratio of the width of the set  $[-2F/s, 2F/s]$  to  $\mathcal{O}(F)$ , see the figure 3.2) over the randomness of  $h_d$ , we have  $\sum_{i=1}^d h_i w'_i \in [-2F/s, 2F/s]$ , which complete the proof.  $\square$



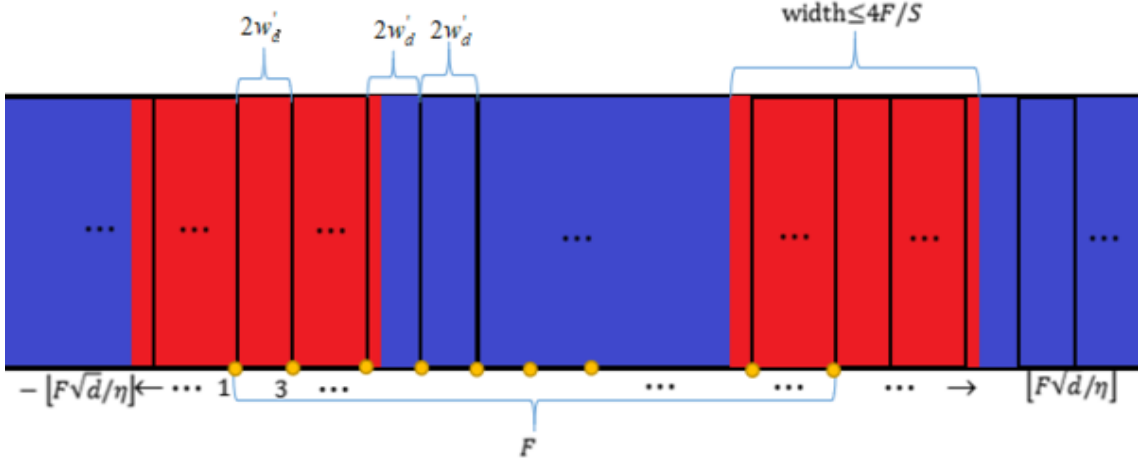


Figure 3.2: Let us assume  $w'_d > \eta/\sqrt{d}$ . Fix arbitrary  $h_1, \dots, h_{d-1}$ . Without loss of generality, suppose  $(h_1, \dots, h_{d-1}, 1)^* \cdot w \in [-2F/s, 2F/s]$ . The red region represents the range of  $h_d$  such that  $\sum_{i=1}^d h_i w'_i \in [-2F/s, 2F/s]$ . As  $h_d$  traverses the set  $\{2n+1 : n \in \mathbb{Z}\} \cap [-\sqrt{d}F/(\eta), \sqrt{d}F/(\eta)]$ , the ratio of the area of the red region to the total area (the sum of the red region and the blue region) is  $\mathcal{O}(1/s)$ , which is approximately equal to the success probability for  $\sum_{i=1}^d h_i w'_i \in [-2F/s, 2F/s]$ .

Notice that the number of vectors  $\{w_i - w_j : 1 \leq i < j \leq k\}$  is  $k(k-1)/2$ . From Lemma 3.2, the probability of the union of the collision events (namely,  $(h(w_i - w_j))_d \in (-2F/s, 2F/s)$ ) is at most

$$\frac{\frac{k(k-1)}{2} \times (\frac{\delta}{8k^2} \times \bar{H})}{\bar{H}} = \frac{k(k-1)\delta}{16k^2},$$

where  $\bar{H}$  denotes the number of elements in the set  $\{2n+1 : n \in \mathbb{Z}\}^d \cap [-\sqrt{d}F/(\eta), \sqrt{d}F/(\eta)]^d$ . Therefore, we have a following consequence.

**Corollary 3.3** *All parameters are set as above lemma. Suppose the vectors  $w_1, \dots, w_k \in [-M, M]^d$  satisfy*

$$\min_{i \neq j} |w_i - w_j| > \eta,$$

*then*

$$\min_{1 \leq i < j \leq k} |(h(w_i - w_j))_d| > 2F/s$$

*holds with probability at least  $1 - \delta/8$  over the randomness of  $h_1, h_2, \dots, h_d$ .*

Suppose  $dF\beta/\eta \leq F\delta/(8ks)$ . For any  $\xi_1, \xi_2 \in \phi_{j,\beta}$ , we have  $|(h(\xi_1) - h(\xi_2))_d| \leq \delta F/(8ks) < F/s$ . However, this does not mean that there is a bucket  $B_{j_s}$  such that  $h(\phi_{j,\beta}) \subset B_{j_s}$ . It is possible that  $h(\phi_{j,\beta})$  intersects on the boundary of some bucket. Obviously, after adding random translation, with a certain probability (the ratio of the total “width” of the sets  $\phi_{j,\beta}, j = 1, 2, \dots, k$  to the “width” of the bucket), there exists a bucket  $B_{j_s}$  such that  $H(\phi_{j,\beta})$  is completely inside it, that is, the following conclusion holds.

**Lemma 3.4** *Suppose  $dF\beta/\eta \leq F\delta/(8ks)$ , with probability at least  $1 - \delta/8$  over the randomness of  $b$ , there exists a bucket  $B_{j_s}$  such that  $H(\phi_{j,\beta}) \subset B_{j_s}$  for  $j=1, 2, \dots, k$ .*



Now we give the proof of Lemma 2.1. Choosing  $s = \mathcal{O}((k^2/\delta))$ ,  $\beta = \eta\delta/(8dks)$ ,  $F = sM$  and  $T = \mathcal{O}(k^5 d^{5/2}/((\frac{\epsilon\sqrt{\delta}}{\sqrt{ds}})^2 \eta \delta^2))$ . By the inequality (3.12), we have

$$\frac{1}{T^d} \sum_{\xi \in \Gamma_2} |\mathcal{X}_{\phi_{j,\beta}}(\xi) \cdot \mathcal{F}[f_j](\xi) - \mathcal{F}(f_j)(\xi)|^2 \leq \mathcal{O}(A^2 \epsilon^2 \delta / (dsk^2)). \quad (3.13)$$

From Corollary 3.3, Lemma 3.4 we know that for each bucket  $B_j$ , there is at most one set  $\phi_{j_k,\beta}$  such that  $H(\phi_{j_k,\beta}) \subset B_j$ . Using the inequality (3.13), since  $\mathcal{X}_{\phi_{j,\beta}}(H^{-1}(\xi)) = \mathcal{X}_{H(\phi_{j,\beta})}(\xi)$ , then with probability at least  $1 - \delta/4$  over the randomness of  $\{h_1, \dots, h_d, b\}$ , for each bucket  $B_j$ , either

$$\frac{1}{T^d} \sum_{\xi \in \Gamma_2} |\mathcal{X}_{B_j}(\xi) \cdot \mathcal{F}[f](H^{-1}(\xi)) - \mathcal{F}(f_{j_k})(H^{-1}(\xi))|^2 \leq \mathcal{O}(A^2 \epsilon^2 \delta / (ds)), \quad (3.14)$$

or

$$\frac{1}{T^d} \sum_{\xi \in \Gamma_2} |\mathcal{X}_{B_j}(\xi) \cdot \mathcal{F}[f](H^{-1}(\xi))|^2 \leq \mathcal{O}(A^2 \epsilon^2 \delta / (ds)). \quad (3.15)$$

Since  $\mathcal{F}[f_H](\xi) = \mathcal{F}[f](H^{-1}\xi)$ , we complete the proof by Parseval's identity.

## 4 Proof of Lemma 4.1

The following lemma reduces the computational complexity of the convolution operation. The classical method to reduce the computational complexity (from  $Poly(TF)$  to  $Poly(\log(TF))$ ) is to select an appropriate window, such as the *Sinc*  $\times$  *Gaussian* window [17], which can approximate the indicative function in the frequency domain and rapidly decay in the time domain. In this paper, we do not add windows for brevity, instead, we choose the appropriate sampling method.

**Lemma 4.1** *For any  $0 < \epsilon < 1/2$ ,  $1 \leq j \leq s$ , suppose*

$$N = \mathcal{O}\left(\frac{k^2 A^2 \ln^2(TF) \ln^2(1/\delta)}{\epsilon^2}\right),$$

*then with probability at least  $1 - \delta/4$  over the randomness  $\mathcal{T}$ , the following inequality holds*

$$\sup_{x \in \Gamma_1} \left| \mathcal{F}^{-1}[\mathcal{X}_{B_j} \cdot \mathcal{F}[f_H]](x) - \frac{1}{N} \sum_{i=1}^N f_H\left(x - \left(0, \dots, 0, \frac{\lfloor \text{sgn}(t_i)((\frac{TF}{2} + 1)^{2|t_i|} - 1) \rfloor}{F}\right)^*\right) v_j(t_i) \right| \leq \epsilon, \quad (4.16)$$

where  $\mathcal{T} \triangleq \{t_1, \dots, t_N\}$  are independent draws from the uniform distribution on  $(-1/2, 1/2)$ ,

$$v_j(t) = 2 \ln\left(\frac{TF}{2} + 1\right) \left(\frac{TF}{2} + 1\right)^{2t} v_{2,j}\left(\left\lfloor \text{sgn}(t) \left(\left(\frac{TF}{2} + 1\right)^{2|t|} - 1\right) \right\rfloor\right),$$

and

$$v_{2,j}(y) \triangleq \begin{cases} \frac{\exp(-\pi i y) (\exp(\frac{2\pi i y(j-1)}{s}) - \exp(\frac{2\pi i y j}{s}))}{TF(1 - \exp(2\pi i y / TF))} & y \neq 0 \\ 1/s & y = 0 \end{cases}. \quad (4.17)$$



*Proof:* Observe that

$$\sum_{w \in \mathbb{Z} \cap [-TF/2, TF/2]} \exp\left(\frac{2\pi i n w}{TF}\right) = 0 \quad \text{for } n \in \mathbb{N}^+ \cup \mathbb{N}^-,$$

then

$$\begin{aligned} \mathcal{F}^{-1}[\mathcal{X}_{B_j} \cdot \mathcal{F}[f_H]](x) &= \sum_{y \in \Gamma_1} \frac{f_H(x-y) \mathcal{F}^{-1}(\mathcal{X}_j)(y)}{(\sqrt{TF})^d} \\ &= \sum_{z \in \mathbb{Z} \cap [-\frac{TF}{2}, \frac{TF}{2}]} f_H(x - (0, \dots, 0, \frac{z}{F})^*) v_{2,j}(z) \\ &= \int_0^{TF/2} f_H(x - (0, \dots, 0, \frac{\lfloor z \rfloor}{F})^*) (z+1) 2 \ln(1 + TF/2) v_{2,j}(z) d\left(\frac{\ln(z+1)}{2 \ln(1+TF/2)}\right) \\ &+ \int_{-TF/2}^0 f_H(x - (0, \dots, 0, \frac{\lfloor z \rfloor}{F})^*) (-z+1) 2 \ln(1 + TF/2) v_{2,j}(z) d\left(\frac{\ln(-z+1)}{2 \ln(1+TF/2)}\right) \\ &= \int_{-1/2}^{1/2} f_H(x - (0, \dots, 0, \frac{\lfloor \text{sgn}(t)((\frac{TF}{2}+1)^{2|t|}-1) \rfloor}{F})^*) v_j(t) dt, \end{aligned} \quad (4.18)$$

where  $\sup_{t \in [-1/2, 1/2]} |v(t)| \leq \mathcal{O}(\ln(TF))$ .

When we fix  $x$ , the above integral can be calculated by Monte Carlo method, and the error can be estimated by Hoeffding's inequality. However, when we want to obtain the uniform error for all  $x$  in  $\Gamma_1$ , we need to use the Rademacher complexity [27, 28]. We consider the Rademacher complexity of the following function spaces

$$Q_{w,u,C_0} \triangleq \{q_x(t) \triangleq u(t) \sin(2\pi w (\frac{x - \lfloor \text{sgn}(t)((\frac{TF}{2}+1)^{2|t|}-1) \rfloor}{F}))\},$$

and

$$Q'_{w,u,C_0} \triangleq \{q'_x(t) \triangleq u(t) \cos(2\pi w (\frac{x - \lfloor \text{sgn}(t)((\frac{TF}{2}+1)^{2|t|}-1) \rfloor}{F}))\},$$

where  $w, x, t \in \mathbb{R}$  and  $u(t)$  is any real-valued function satisfying  $|u| \leq C_0$ . Let  $\widehat{\text{Re}}(\mathcal{T}; Q_{w,u,C_0})$  denote the empirical Rademacher complexity for  $Q_{w,u,C_0}$  and  $\mathcal{T}$ , then

$$\begin{aligned} \widehat{\text{Re}}(\mathcal{T}; Q_{w,u,C_0}) &= \mathbb{E}_{\xi \sim \{\pm 1\}^N} \left[ \sup_{x \in \mathbb{R}} \sum_{i=1}^N \frac{\xi_i q_x(t_i)}{N} \right] \\ &\leq \mathbb{E}_{\xi \sim \{\pm 1\}^N} \left[ \sup_{(z_1, z_2) \in [-1, 1]^2} \sum_{i=1}^N \frac{\xi_i (z_1 y_{i,1} - z_2 y_{i,2})}{N} \right] \leq \frac{2C_0}{\sqrt{N}}, \end{aligned} \quad (4.19)$$

where

$$y_{i,1} = \cos\left(\frac{2\pi w \lfloor \text{sgn}(t_i)((\frac{TF}{2}+1)^{2|t_i|}-1) \rfloor}{F}\right) u(t_i), \quad y_{i,2} = \sin\left(\frac{2\pi w \lfloor \text{sgn}(t_i)((\frac{TF}{2}+1)^{2|t_i|}-1) \rfloor}{F}\right) u(t_i),$$

the last inequality in (4.19) is obtained by Lemma 26.10 in [27]. Similarly, we have  $\widehat{\text{Re}}(\mathcal{T}; Q'_{w,u,C_0}) \leq \frac{2C_0}{\sqrt{N}}$ .

Combining the last equality in (4.18) and Lemma A.10 in [28] proves Lemma 4.1.  $\square$

## 5 Conclusion

The hashing transform we construct is also applicable to the case of lattice (as considered in [20, 21, 22, 23]), which we have not covered in this paper. We do not consider the case of noise, so it is an issue



that need to be discussed later. In addition, whether the high-dimensional signal without frequency gap can be effectively reconstructed (just like one-dimensional case [25])? In terms of application, it is worthy to optimize the algorithm and the complexity estimation to make it more applicable to practice.

## References

- [1] A. C. Gilbert, P. Indyk, M. Iwen, and L. Schmidt, “Recent developments in the sparse fourier transform: A compressed fourier transform for big data,” *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 91–100, 2014.
- [2] E. Kushilevitz and Y. Mansour, “Learning decision trees using the fourier spectrum,” *SIAM Journal on Computing*, vol. 22, no. 6, pp. 1331–1348, 1993.
- [3] M. A. Iwen, “Combinatorial sublinear-time fourier algorithms,” *Foundations of Computational Mathematics*, pp. 303–338, 2010.
- [4] —, “Improved approximation guarantees for sublinear-time fourier algorithms,” *Applied And Computational Harmonic Analysis*, vol. 34, no. 1, pp. 57–82, 2013.
- [5] D. Potts and T. Volkmer, “Sparse high-dimensional fft based on rank-1 lattice sampling,” *Applied and Computational Harmonic Analysis*, vol. 41, no. 3, pp. 713–748, 2016.
- [6] S. Merhi, R. Zhang, M. A. Iwen, and A. Christlieb, “A new class of fully discrete sparse fourier transforms: Faster stable implementations with guarantees,” *Journal of Fourier Analysis and Applications*, no. 1, pp. 1–34, 2017.
- [7] S. Bittens, R. Zhang, and M. A. Iwen, “A deterministic sparse fft for functions with structured fourier sparsity,” *Advances in Computational Mathematics*, vol. 45, no. 2, pp. 519–561, 2019.
- [8] B. Choi, M. A. Iwen, and F. Krahmer, “Sparse harmonic transforms: A new class of sublinear-time algorithms for learning functions of many variables,” *Foundations of Computational Mathematics*, pp. 1–55, 2020.
- [9] S.-H. Hsieh, C.-S. Lu, and S.-C. Pei, “Sparse fast fourier transform by downsampling,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 5637–5641.
- [10] S. Liu, T. Shan, R. Tao, Y. D. Zhang, and Y. Wang, “Sparse discrete fractional fourier transform and its applications,” *IEEE Transactions on Signal Processing*, vol. 62, no. 24, p. 6582C6595, 2014.
- [11] L. Shi, H. Hassanieh, A. Davis, D. Katabi, F. Durand. “Light field reconstruction using sparsity in the continuous fourier domain,” *ACM Transactions on Graphics*, pp. 1–13, 2014.
- [12] S. Pawar and K. Ramchandran, “FFAST: An algorithm for computing an exactly  $k$ -sparse DFT in  $\mathcal{O}(k \log k)$  time,” *IEEE Transactions on Information Theory*, vol. 64, no. 1, pp. 429–450, 2017.
- [13] S. Wang, V. M. Patel, and A. Petropulu, “The robust sparse fourier transform (rsft) and its application in radar signal processing,” *IEEE Transactions on Aerospace and Electronic Systems*, pp. 2735–2755, 2017.
- [14] —, “Multidimensional sparse fourier transform based on the fourier projection-slice theorem,” *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 54–69, 2018.
- [15] A. C. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. Strauss, “Near-optimal sparse fourier representations via sampling,” in *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, 2002, pp. 152–161.
- [16] A. C. Gilbert, S. Muthukrishnan, and M. Strauss, “Improved time bounds for near-optimal sparse fourier representations,” in *Wavelets XI*, vol. 5914. International Society for Optics and Photonics, 2005, p. 59141A.



- [17] H. Hassanieh, P. Indyk, D. Katabi, and E. Price, “Simple and practical algorithm for sparse fourier transform,” in *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*. SIAM, 2012, pp. 1183–1194.
- [18] P. Boufounos, V. Cevher, A. C. Gilbert, Y. Li, and M. J. Strauss, “Whats the frequency, kenneth?: Sublinear fourier sampling off the grid,” *Algorithmica*, vol. 73, no. 2, pp. 261–288, 2015.
- [19] M. Kapralov, “Sparse fourier transform in any constant dimension with nearly-optimal sample complexity in sublinear time,” in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, 2016, pp. 264–277.
- [20] M. Kapralov, A. Velingker, and A. Zandieh, “Dimension-independent sparse fourier transform,” in *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2019, pp. 2709–2728.
- [21] L. Kämmerer, D. Potts, and T. Volkmer, “High-dimensional sparse FFT based on sampling along multiple rank-1 lattices,” *Applied and Computational Harmonic Analysis*, vol. 51, pp. 225–257, 2020.
- [22] B. Choi, A. Christlieb, and Y. Wang, “Multiscale high-dimensional sparse fourier algorithms for noisy data,” *arXiv preprint arXiv:1907.03692*, 2019.
- [23] —, “High-dimensional sparse fourier algorithms,” *Numerical Algorithms*, pp. 1–26, 2020.
- [24] E. Price and Z. Song, “A robust sparse fourier transform in the continuous setting,” in *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*. IEEE, 2015, pp. 583–600.
- [25] X. Chen, D. M. Kane, E. Price, and Z. Song, “Fourier-sparse interpolation without a frequency gap,” in *2016 IEEE 57th Annual Symposium on Foundations of Computer Science*. IEEE, 2016, pp. 741–750.
- [26] Y. Jin, D. Liu, and Z. Song, “A robust multi-dimensional sparse fourier transform in the continuous setting,” *arXiv preprint arXiv:2005.06156*, 2020.
- [27] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
- [28] Z. Allen-Zhu, Y. Li, and Y. Liang, “Learning and generalization in overparameterized neural networks, going beyond two layers,” in *Advances in neural information processing systems*, 2019, pp. 6158–6169.